

# Do Rewards Encourage Professional Skepticism?

## Abstract

It is an open question whether auditors have credible incentives to exercise professional skepticism. We focus on costly skepticism: skepticism that is appropriate and generates incremental costs, but does not identify a misstatement. Costly skepticism is typically not rewarded by audit supervisors. We theorize and find that rewarding costly skepticism may backfire and decrease skepticism on subsequent audit tasks where evidential red flags are present. We reason that auditors interpret the reward as a non-credible, better-than-expected outcome, leading auditors to view subsequent tasks from a risk-averse gain frame. As a result, auditors self-interestedly seek to avoid the risks and effort of exercising additional skepticism. This effect decreases auditors' sensitivity to red flags and auditors' willingness to inform their manager about severe red flags, compromising audit quality. Encouragingly, auditors who have experienced a history of rewards for costly skepticism are more motivated to exercise skepticism. A survey finds that audit supervisors are likely to reward costly skepticism when their own supervisors encourage the behavior and promote consultation within the engagement team. Overall, our results suggest firms may benefit from a culture shift emphasizing credible rewards for costly skepticism, but that firms currently may not "get what they reward."

**Keywords:** performance evaluation, professional skepticism, motivation, strategic behavior

**Data Availability:** *Contact the authors.*

## 1. Introduction

Professional skepticism is essential to audit quality, and enhancing auditor skepticism is of great concern to regulators, practitioners, and scholars (e.g., Nelson [2009]; PCAOB [2012]; IAASB [2015]; KPMG [2016]).<sup>1</sup> However, little is known about the conditions in which individual auditors have sufficient incentives to exercise skepticism, and it may be difficult for firms to provide these incentives. Pervasive fee pressures lead to tight budgets, and audit firms often absorb the costs of overruns (Dopuch et al. [2003]; Hackenbrack and Hogan [2006]). As a result, auditor performance evaluations emphasize meeting time budgets (Agoglia et al. [2015]), which may discourage skeptical behavior. Moreover, skepticism that is *ex ante* appropriate often *does not* identify a misstatement, but *does* typically generate incremental *ex post* costs (e.g., budget overruns or strained client relations). We use the term “costly skepticism” to describe appropriate applications of skepticism that incur costs, yet do not ultimately identify a misstatement.<sup>2</sup> Costly skepticism carries substantial downside risk for individual auditors, as supervisors provide *disincentives* for costly skepticism even when they acknowledge its appropriateness (Brazel et al. [2016]). Thus, auditors may not have adequate incentives to exercise skepticism.

An intuitive solution is to implement rewards for costly skepticism, as economic theory argues that rewarding a behavior should encourage that behavior (e.g., Prendergast [1999]). However, we argue that rewards may not increase skepticism due to two opposing effects. First, rewarding costly skepticism may trigger self-interested behavior that *decreases* skepticism on

---

<sup>1</sup> Professional skepticism is an attitude that includes a questioning mind and critical assessment of audit evidence, and is demonstrated through skeptical actions (e.g., IAASB [2004]; Nelson [2009]; PCAOB [2012]).

<sup>2</sup> To be clear, skepticism that yields a misstatement is also likely to incur costs. However, Brazel et al. [2016] and Brazel et al. [2018] demonstrate that such skepticism is likely to be viewed by superiors as a *normal* cost of the audit that benefits the engagement team. These studies also demonstrate that evaluators reliably reward skepticism that identifies a misstatement. In subsequent analyses we illustrate that while reliably rewarding *costly* skepticism has the potential to enhance auditor skepticism, reliable rewards for exercising skepticism that yields a misstatement does not.

subsequent tasks. Auditors know that rewards for costly skepticism are uncommon, and thus may view rewards as non-credible and unlikely to repeat (Brazel et al. [2016]; Brazel et al. [2018]). Instead, a reward is likely a better-than-expected outcome that may cause auditors to view subsequent tasks from a risk-averse *gain frame* (e.g., Kahneman and Tversky [1979]). This leads to a self-interested “quit while you’re ahead” mentality, which could prompt auditors to avoid the risks of further skeptical behavior – ultimately *decreasing* skepticism. At the same time, auditors may also interpret a reward as a signal that costly skepticism is a desirable behavior, which can increase motivation and subsequent skepticism (e.g., Bonner et al. [1997]). Because these two causal mechanisms have opposing signs, the adage “you get what you reward” may not apply to rewarding skepticism.

We also examine the severity of red flags or inconsistencies in the evidence available to auditors as a potential moderator of the aforementioned effects. Auditors confront increasingly complex information with some data points that support client assertions and other data that are inconsistent with client assertions. Appropriate responses to inconsistent information are crucial to audit quality (Griffith et al. [2015]), and thus appropriate skepticism requires that auditors exercise heightened skepticism as the severity of evidential red flags increase (Nelson [2009]). We predict that as the severity of red flags increases, the aforementioned skepticism-decreasing self-interest effect will decrease, and the skepticism-increasing motivation effect will increase.

We tested our hypotheses in an experiment with 112 audit seniors. At the onset of the experimental case, auditors were informed that this was their first year serving on the engagement (i.e., they had no prior year experience with the supervisor) and that they had engaged in costly skepticism during interim testing. Half of the auditors were rewarded for this work with a positive evaluation, whereas the other half received a middling evaluation and thus

were not rewarded. The case then asked the auditors to perform an analytical procedure for sales based on a rich information set of financial data and non-financial measures (NFMs). We held sales growth constant, and the associated financial data (e.g., industry data) were consistent with the level of sales growth. However, NFM growth for the client was negative in all cases, creating a red flag. Our primary measure of skepticism was whether or not auditors concluded that the sales account warranted additional investigation. We manipulated the trend in NFMs (e.g., number of customers) to be increasingly negative, specifically as a minor, moderate, or severe red flag (Brazel et al. [2009]). This enabled us to test the *appropriateness* of auditor skepticism (i.e., heightened skepticism as the red flag becomes more severe).

Our predictions are two opposite-signed indirect effects, with rewards decreasing skepticism due to higher auditor self-interest and increasing skepticism due to higher auditor motivation. Consistent with auditor self-interest, we find that auditors interpret a reward (vs. no reward) for costly skepticism as a better-than-expected outcome and personal benefit, which in turn decreases their subsequent skepticism. Contrary to expectations, we do not find a skepticism-increasing motivation effect for rewards. Supplemental tests support our theory that auditors believe rewards for costly skepticism are infrequent and unlikely to repeat. Auditors who were rewarded for costly skepticism did not expect another reward for their work on the experimental task, regardless of the skepticism they exhibited.

Further, we find no evidence that red flag severity moderates either the self-interest or motivation effects in our experiment. Instead, auditors rewarded for costly skepticism *do not* increase skepticism towards *more severe* red flags, thus failing to act skeptically when doing so is most appropriate. Additional tests reveal that auditors believe the costs of skepticism would be quite high in this condition, possibly explaining their reluctance to act skeptically. Moreover,

auditors rewarded for costly skepticism recognize that increasingly severe red flags are inconsistent with client assertions (i.e., they exercise appropriate skeptical *judgment*), but choose not to *act* on their judgment. Being aware of, but not appropriately considering, contradictory audit evidence is a behavior consistent with PCAOB inspection findings (PCAOB 2012). These auditors are also less willing to communicate the severe red flag to their manager, further compromising audit quality by undermining a basic quality control (Nelson et al. [2016]; Lightle et al. [2017]). In sum, our evidence suggests that audit firm evaluation cultures may fail to provide credible incentives for skepticism.

Examining the indirect effects of rewards on skepticism allows us to better understand the causal mechanisms in the relation between rewards and the application of skepticism, which is essential to improving incentives for skepticism. These dysfunctional incentives apply to a range of common settings in which auditors work with relatively unfamiliar supervisors (e.g., all first year assignments, recurring assignments on which supervisors change due to turnover or promotion, experienced supervisor hires, firm human resource needs that assign junior or supervising auditors to a new client, industry, or office). Indeed, Christensen and Newton [2017] provide archival evidence on engagement team turnover, and illustrate that approximately half of the hours charged by team members are charged by members who were not assigned to the engagement in the prior year.

Still, it is possible that supervisors can develop reputations for *reliably* rewarding costly skepticism. Auditors who have experience with such supervisors may interpret a reward for costly skepticism as a credible signal, increasing their willingness to exercise skepticism on subsequent tasks, as they believe the reward will be repeated. To examine this possibility, we measured participants' personal experiences being rewarded by their supervisors for costly

skepticism. Encouragingly, we find that auditors with a history of being rewarded have higher motivation to exercise professional skepticism on subsequent tasks. However, this experience does not nullify the self-interest effect of rewards that decreases skepticism. Given the prevalence of supervisors who do not reward costly skepticism, auditors most likely still consider their own best interest and avoid skeptical action if they are unsure about the supervisor. Overall, our results highlight the necessity for a culture shift where rewarding costly skepticism is the norm, not the exception, so that rewards may begin to have the intended effect.

Finally, because rewards for costly skepticism can be effective if auditors believe the rewards to be credible, we conducted a case-based survey of 127 audit supervisors to explore attributes of supervisors *who may credibly reward costly skepticism*. In the survey, supervisors evaluated the performance of a subordinate who engaged in appropriate, yet costly skepticism. We find supervisors' evaluations increase with (1) beliefs that their own superiors would reward them for a subordinate's costly skepticism, and (2) more experience in consultative work environments. Note that both involve supportive superiors and positive experiences over time, suggesting professional skepticism is a nurtured, learned behavior. This is consistent both with our experimental findings that a supportive culture can motivate professional skepticism and with regulators' emphasis on "tone at the top" vis-à-vis professional skepticism (PCAOB [2012]). Our results should also inform standard setters' current deliberations over audit supervision and other quality controls to foster appropriate skepticism (PCAOB [2017]; IAASB [2017]).<sup>3</sup>

---

<sup>3</sup> In addition, the topic of professional skepticism was discussed at the most recent PCAOB Standing Advisory Group Meeting (see <https://pcaobus.org/News/Events/Pages/SAG-meeting-Nov-2017.aspx>), as well as several recent IAASB Meetings (see [https://www.iaasb.org/system/files/meetings/files/20160314-IAASB-Agenda\\_Item\\_2A-Professional\\_-Skepticism-Discussion-final\\_1.pdf](https://www.iaasb.org/system/files/meetings/files/20160314-IAASB-Agenda_Item_2A-Professional_-Skepticism-Discussion-final_1.pdf)).

As a whole, our findings suggest that auditor performance evaluation systems prompt dysfunctional responses from auditors, creating a vicious cycle in which positive evaluations of costly skepticism can discourage skepticism on subsequent tasks. This cycle will likely become more crucial to audit quality as data analytics and identifying outliers play a larger role in the audit process (KPMG [2017]; EY [2017]). However, our results also point to a solution: a culture shift towards reliably rewarding appropriate skepticism, regardless of the outcome. Subordinates who regularly experience such rewards are more motivated to engage in skepticism. Further, supervisors reward costly skepticism when they experience consultative and supportive environments in which they expect to be rewarded for their subordinates' skepticism. As a result, a culture supportive of costly skepticism and consultation may create a virtuous cycle in which supervisors encourage skepticism and subordinates exhibit skepticism.

Although audit firm incentive systems likely function well in many conditions, we argue that these systems do not align individual auditor incentives with the broader goal of audit quality on the crucially important dimension of professional skepticism. It is critical for audit firms to craft quality control systems, mentoring programs, and trainings to promote credible rewards for costly skepticism. For example, firms routinely train auditors to engage in skepticism, but with cases that involve misstatements. Refining this approach would acclimate junior auditors and supervisors to exercising and rewarding *costly* skepticism.

## *2. Theory and Hypothesis Development*

### **2.1 PERFORMANCE EVALUATION AND PROFESSIONAL SKEPTICISM**

Exercising professional skepticism is fundamental to auditing and increases audit quality (e.g., PCAOB [2012]; IAASB [2012]; IAASB [2015]; KPMG [2016]). However, regulators consistently note insufficient skepticism on audit engagements and question the degree to which

firms' systems of quality control encourage skepticism (e.g., PCAOB [2012]; IAASB [2015]). A firm's performance evaluation and incentive system is a key quality control component that theory suggests should reward employee actions that align with firm interests – namely those contributing to efficient and effective audits (e.g., Baiman [1990]; IAASB [2009]). Thus, effective evaluation and reward systems should enable firms to address inadequate skepticism. Little is known, however, about how such rewards currently affect auditor skepticism.

In auditing, as in other professions, evaluations are subjectively-assessed composites capturing objective outcomes and a variety of hard-to-measure signals relevant to the employee's contribution to the firm (Werner [1994]; Van Scotter et al. [2000]; Witt and Ferris [2003]). The primary objective measures on which supervisors evaluate auditors relate to chargeable hours and adherence to time budgets (Agoglia et al. [2015]). Supervisors also evaluate subjective dimensions, such as the appropriate application of skepticism, client relations, judgment quality, technical knowledge, and teamwork. It is particularly difficult for supervisors to evaluate the appropriate application of skepticism, because skepticism can conflict with objective indicators. For instance, exercising skepticism may be necessary in the presence of red flags, but doing so can lead to poor objective performance in the form of budget overruns (Houston [1999]; Nelson [2009]; Peecher et al. [2013]).

Difficulty evaluating skepticism leads to difficulty in providing incentives for skepticism, which is further complicated by the low base rate of materially misstated accounts.<sup>4</sup> Auditors often confront a *false positive paradox* in which cues suggestive of a misstatement (e.g., red flags) are more likely to reflect non-misstatements than misstatements (Waller and Zimbelman

---

<sup>4</sup> Evidence suggests the base rate of materially misstated accounts is quite low. For example, Durney et al. [2014] examine 160 sampling applications across 51 audit engagements and find that nearly 60% contain zero misstatement and 90% are misstated by 0.5% or less of the recorded balance. Consistent with this notion, 26% of the auditors in our experiment reported that they had *never* detected a material misstatement.



[2003]). Nevertheless, auditing standards require auditors to maintain a questioning mindset and respond to cues that may signal a misstatement (IAASB [2004]; PCAOB [2015]; [2016]). Thus, skepticism is often simultaneously appropriate, yet directed towards fairly-stated accounts. We refer to appropriate skeptical behavior that generates costs, but does not yield a misstatement as *costly skepticism*. Because audit supervisors respond negatively to budget overruns, appropriate skepticism can lead directly to poor performance on the primary objective measure that determines the auditor's evaluation (e.g., Agoglia et al. [2015]).

In an effective reward system, supervisors would offset this downside risk for the auditor with a positive subjective evaluation for the appropriate application of skepticism. There is evidence that supervisors do so in many settings (Gibbs et al. [2004]; Bol and Smith [2011]). However, Brazel et al. [2016] provide evidence that audit supervisors discourage costly skepticism, even when they acknowledge that the skepticism is appropriate. Moreover, Brazel et al. [2016] illustrate that junior auditors *know* that supervisors will react negatively to costly skepticism, which further limits auditors' willingness to exercise skepticism.

Performance evaluation theory suggests an intuitive solution to this issue: change evaluation systems so that supervisors offset poor objective performance with positive subjective evaluations of costly skepticism (e.g., McLeod [2003]; Bol [2008]). Decades of research demonstrates that rewarding a behavior is likely to motivate more of that behavior (Prendergast [1999]). If auditors trust that supervisors will reward costly skepticism, then the prospect of positive evaluations should generate more skepticism. However, because subjective evaluations in auditing are often biased and auditors are aware of the bias (Tan and Jamal [2001]; Brazel et al. [2016]), positive evaluations may not result in the expected increase in skepticism. In fact, rewarding costly skepticism may actually *discourage* skepticism on subsequent audit tasks.

## 2.2 HYPOTHESIS DEVELOPMENT

To understand why rewarding costly skepticism may *discourage* subsequent skepticism rather than *encourage* it, one must consider auditor perceptions of evaluation norms in audit firms. Auditors justifiably believe that rewards for costly skepticism are uncommon (Brazel et al. [2016]; Brazel et al. [2018]).<sup>5</sup> All else equal, auditors also likely understand that receiving such a reward is a benefit that few of their peers will receive. This is significant because audit firms use relative evaluation systems in which small absolute differences in performance between peers can lead to large differences in overall rewards (e.g., Holmstrom [1981]; Cichello et al. [2009]).<sup>6</sup> Prior literature demonstrates that junior auditors are effective strategists and recognize the actions that serve their self-interest (Rich et al. [1997]; Mayhew [2001]; Tan and Jamal [2006]). Consistent with these findings, auditors may respond to rewards for costly skepticism in a strategic, self-interested manner.

Specifically, we argue that benefitting from an act of costly skepticism is a better-than-expected outcome leading auditors to view subsequent tasks from a gain frame (e.g., Kahneman and Tversky [1979]). A gain frame likely focuses auditors on avoiding the risks and effort of skeptical behavior on their next task, focusing instead on maximizing personal benefits. Building on this idea, there is evidence that reward systems exhibit regression towards the mean—that is, an evaluation outcome that is far from the norm is likely to be followed by an evaluation outcome closer to the norm (Kahneman and Tversky [1973]). In addition, people view evaluation

---

<sup>5</sup> Altogether, Brazel et al. [2016] and Brazel et al. [2018] observe that costly skepticism is not rewarded even in the following contexts: when the skeptical auditor consults and obtains their superior's approval prior to engaging in skeptical actions, when the skeptical auditor evidences his skepticism in the budget file as justification for the budget overage, and when the audit committee is likely to insulate the audit team from the costs of skepticism (budget overruns and strained management relations).

<sup>6</sup> Auditors typically work on multiple engagement teams, receiving evaluations from multiple supervisors and often receiving multiple evaluations for the same engagement (Capelli and Tavis [2016]; Feintzeig [2017]; Reitman [2017]). Audit firms combine the various evaluations for a given auditor into a single, yearly evaluation and then rank-order the auditor against others at the same experience level. Firms then use these rankings to sort auditors into groups to determine rewards such as raises, bonuses, promotions, client assignments, etc.

outcomes that seem far from the norm with suspicion and respond with lower productivity (Berger et al. [2013]). This suggests that auditors are unlikely to believe that a reward for costly skepticism will be followed by another reward for costly skepticism. Instead, a reward for costly skepticism may instead reinforce that auditors can benefit themselves by avoiding the risks and costs of skeptical action. This is consistent with evidence that implementing rewards can unintentionally highlight the benefits of self-interested behavior (Cardinaels and Yin [2015]). Thus, auditors rewarded for costly skepticism are likely to adopt a “quit while you’re ahead” mindset that decreases subsequent skepticism.

Skeptical behavior risks poor performance on at least one key objective performance dimension: going over budget. Given the low probability of either detecting a misstatement or receiving another reward for costly skepticism, the cautious route is *not* to exercise skepticism. Doing so offers a high probability of good performance on the objective time budget measure, and the low base rate of misstatements means that the likelihood of missing a misstatement is relatively low. That is, auditors have more to lose than to gain from skeptical action. Moreover, auditors are likely able to rationalize diminished skepticism, as they commonly approach tasks expecting no misstatement to exist and are skilled at developing non-misstatement explanations for red flags (Kaplan and Reckers [1989]; Kaplan et al. [1992]; Solomon et al. [1999]; Earley [2001]).<sup>7</sup>

Thus, an auditor rewarded for costly skepticism will recognize the reward as a benefit that increases their relative ranking, but also interpret the reward as non-credible and unlikely to repeat. As a result, the auditor may self-interestedly become less willing to engage in skeptical behavior. This leads to our first hypothesis, stated formally:

---

<sup>7</sup> Auditors prefer to maintain a self-image as a good professional (Kadous et al. [2003]), and the reward helps legitimize this self-perception. In general, positive responses to costly behavior often allow people to rationalize not engaging in that behavior in the future (Blanken et al. [2015]).

*H1A:* A supervisor rewarding costly skepticism increases the auditor's perceived personal benefits and, in turn, decreases the auditor's subsequent skepticism on that engagement.

Of course, auditors may simultaneously interpret a reward as a signal that applying skepticism is a valuable component of a high quality audit. Rewards for a behavior tend to increase expectations that a behavior will have positive outcomes for the employee (Bonner and Sprinkle [2002]). As such, a reward for costly skepticism could increase motivation to engage in subsequent skepticism. Similar to the adage "you get what you reward," we expect auditors who receive rewards for costly skepticism to be more motivated to exercise skepticism on subsequent audit tasks. This motivation should increase auditor skepticism on subsequent tasks. This leads to our second hypothesis, stated formally:

*H1B:* A supervisor rewarding costly skepticism increases the auditor's motivation to be skeptical and, in turn, increases the auditor's subsequent skepticism on the engagement.

## 2.3 RED FLAG SEVERITY AND THE APPROPRIATENESS OF SKEPTICISM

Nelson [2009] observes that appropriate skepticism is conditional on the evidence set observable to the auditor. Such evidence includes "exceptions" that represent unusual patterns or noteworthy deviations indicative of heightened misstatement risk. Skepticism is more appropriate when targeted towards more severe red flags, as the appropriate level of skepticism is a function of assessed risks (Glover and Prawitt [2014]).

H1A predicts that auditors will recognize the relative benefits of higher evaluations and respond strategically, decreasing skepticism on subsequent audit tasks. We expect that the presence of more severe red flags will mitigate this effect. While auditors can stylize workpapers to persuade a supervisor that their work was adequate (Rich et al. [1997]; Tan and Yip-Ow [2001]), severe red flags are more likely to catch a supervisor's attention, increasing the potential

costs of inadequate skepticism for the subordinate. Not investigating more severe or obvious red flags also heightens the engagement team's exposure to negative inspections findings, litigation losses, and other ex-post reviews which are likely to view the evidence set observed by the auditor with hindsight bias.<sup>8</sup> Although auditors become skilled at developing non-misstatement explanations as they gain experience and specialized knowledge (Solomon et al. [1999]; Shaub and Lawrence [1999]), junior auditors may lack the skill to do so for severe red flags. As a result, as the severity of the red flag increases, we expect that even self-interested auditors may not be able to "explain away" or ignore the red flag. Thus, we expect the negative effect of rewards for costly skepticism to be mitigated as red flags become more severe.

By contrast, if motivation increases due to a reward (as predicted in H1B), then we expect this effect to be amplified by increasingly severe red flags. In general, increased motivation should improve performance in subsequent audit tasks, especially those in which performance is effort-sensitive (Kennedy [1993]; Libby and Luft [1993]). As a result, to the extent that rewards for costly skepticism increase the motivation to engage in skepticism, auditors will be better able to identify and differentiate between lower and higher risk items, increasing skepticism in the presence of severe red flags. This leads to our second hypothesis:

*H2:* More severe red flags are likely to decrease the personal benefit effect of rewards and increase the motivational effect of rewards.

### *3. Auditor Experiment*

#### **3.1 PARTICIPANTS**

The participants in our experiment were 112 audit seniors from an international accounting firm who completed the experiment during a firm-sponsored training session. Their task was a substantive analytical procedure for sales, which is appropriate for audit seniors

---

<sup>8</sup> <http://ww2.cfo.com/auditing/2014/02/pcaobs-audit-quality-highly-suspect>.

(Trompeter and Wright [2010]; Brazel et al. [2014]). The mean task completion time was 34 minutes. On average, participants had 34 months of experience and conducted analytical procedures related to sales four times during their careers. Two participants did not complete one of the measures necessary for our hypothesis tests, thus our final sample was 110 auditors.

Our experiment was a 2 (Reward: present, absent) X 3 (Red Flag: minor, moderate, severe) between-participants design. We randomly assigned participants to one of the six experimental conditions. We describe the variables in detail below.

### 3.2 DESCRIPTION OF THE EXPERIMENTAL CONTEXT

The experimental materials first informed participants that they were the audit senior conducting year-end procedures for the audit of Madison, Inc., a publicly traded manufacturing company. This was the participant's first year serving on the Madison engagement, thus there was no prior evaluation history on the engagement or with the audit manager.<sup>9</sup> Participants then read that they had engaged in costly skepticism while conducting interim procedures. Specifically, while recalculating depreciation expense for the audit of PP&E, they noted year-to-year changes in the client's assumptions for the useful lives of several asset classes. Although the client provided supporting internal documentation, participants had deemed it appropriate to exercise heightened skepticism and conducted additional procedures to support the changes. However, the additional procedures identified no misstatement, caused friction with the client, and caused the audit of PP&E to go over budget. The case then provided participants with their manager's evaluation of their performance for this interim testing (see below for the description of the *REWARD* manipulation).

---

<sup>9</sup> The characteristics of our setting are similar to a range of common settings in which auditors work with relatively unfamiliar supervisors (e.g., all first year assignments, recurring assignments on which supervisors change due to turnover or promotion, experienced supervisor hires, firm human resource needs that assign junior or supervising auditors to a new client, industry, or office).

Next, participants completed a substantive analytical procedure related to the sales account for one of Madison's operating units, Madison Sporting Goods, which reported positive sales growth. The case informed participants that additional testing of sales would likely go over budget and that client management would react negatively to unwarranted additional tests. The case provided participants with a rich set of financial and non-financial data to conduct the analytic. All *financial* measure trends were positive and consistent with the client's sales (e.g., industry data, prior year balances, ratios, budgets). The trend for all *non-financial* measures (NFM), such as the number of customer accounts, patents, and employees, was negative and inconsistent with reported sales. We manipulated between-participants the degree to which this NFM trend was negative (see below for the description of the *RED FLAG* manipulation).

### 3.3 DEPENDENT VARIABLE

After calculating an expectation for sales and computing the difference between their expectation and reported sales, participants provided the study's primary dependent variable: *PROFESSIONAL SKEPTICISM (PS)*. Specifically, participants chose one of three conclusions about whether or not the sales account warranted additional investigation: (1) "The difference is IMMATERIAL and the balance appears reasonable," (2) "The difference is IMMATERIAL, but additional work would be required related to this analytical procedure before concluding the balance appears reasonable," or (3) "The difference is MATERIAL and more work is required related to this analytical procedure."<sup>10</sup>

We coded *PS* as "1" if participants concluded that more work would be required to conclude the balance was reasonable (i.e., options (2) or (3) above), and "0" if participants concluded the balance appeared reasonable and required no additional work (i.e., option (1)

---

<sup>10</sup> We are thankful for the audit partners who reviewed our experimental instrument for providing us with these three outcomes that reflect how differences related to analytical procedures are addressed in practice.

above).<sup>11</sup> Coding responses with additional work as indicative of skepticism is consistent with the professional skepticism literature that stresses the importance of evaluating inconsistent evidence (e.g., IAASB [2004]; Nelson [2009]).

### 3.4 SUPERVISOR EVALUATION REWARDING COSTLY SKEPTICISM (H1A AND H1B)

In this study, we test dual mediation hypotheses. Hypothesis 1A predicts that a reward for costly skepticism (*REWARD*) increases auditors' beliefs about their own personal benefit (*BENEFIT*) and, in turn, indirectly decreases *PS*. Hypothesis 1B predicts that a reward for costly skepticism (*REWARD*) increases auditors' motivation to engage in skeptical behavior (*MOTIVATION*) and, as such, indirectly increases *PS*. We manipulated *REWARD* after participants read the description of their costly skepticism related to PP&E testing at interim (i.e., *before* participants performed the year-end analytical procedure described above). Participants were randomly assigned to an interim evaluation of either "Met Expectations" (*NO REWARD*) or "Exceeded Expectations" (*REWARD*) from their supervisor.<sup>12</sup>

Consistent with Brazel et al. (2016), we measured our first mediator *BENEFIT* with the question "How would you rank in your class (compared to other seniors in your office) if you consistently received the above evaluation?" Participants responded on a scale ranging from 1 (Bottom of My Class) to 10 (Top of My Class). We measured our second mediator *MOTIVATION* with the question "To what extent were you motivated to exercise professional skepticism when auditing the 12/31/12 Madison sporting goods sales account?" Participants responded on a scale ranging from 1 (Not at all Motivated) to 10 (Extremely Motivated).

---

<sup>11</sup> We dichotomized this variable because both options (2) and (3) involve further investigation (i.e., skeptical action) and because only nine participants chose option (3). Our inferences are the same if we use a dependent variable with three levels instead of two.

<sup>12</sup> The experimental materials made it clear that the *REWARD/NO REWARD* interim evaluation was for costly *PS* and there was no other performance-related information presented. See Brazel et al. (2016) for a discussion of how auditors do not consider an evaluation of "Met Expectations" as a reward.



### 3.5 APPROPRIATENESS OF SKEPTICISM (H2) – SEVERTY OF RED FLAGS

To examine the appropriateness of *PS*, we manipulated the severity of the NFM *RED FLAG* (*RF*) in the case. Sales growth was constant at 9% in all conditions, and all conditions included NFMs with negative growth, inconsistent with reported sales. We manipulated the NFM trend to create three levels of *RF*: *MINOR*, *MODERATE*, and *SEVERE*. NFM growth for the *MINOR*, *MODERATE*, and *SEVERE* conditions was -1%, -21%, and -41%, respectively. See the Appendix for a description of all variables used in this study.

### 3.6 PRIMARY RESULTS

#### 3.6.1. Manipulation Checks

For *REWARD*, a post-experimental question asked participants whether they received a “Met Expectations” or “Exceeded Expectations” evaluation for their interim work, and 95 of the 110 participants (86%) answered correctly.<sup>13</sup> As a manipulation check for *RF*, we asked participants to assess the trend in NFMs from 1 (Very Negative) to 10 (Very Positive). Consistent with an effective manipulation, *RF* significantly affects NFM trend assessments ( $F_{2, 103} = 17.05$ ;  $p < 0.01$ ), and the assessments decrease monotonically from the minor to moderate to severe conditions.

#### 3.6.2. Univariate Statistics

Table 1, Panel A reports univariate values of *BENEFIT*, *MOTIVATION*, and *PS* across *REWARD* conditions. Panel B reports *PS* values broken out across *RF* conditions. We do not observe a univariate difference in *PS* across *REWARD* conditions, which is logical given that we predict opposite-signed indirect effects that could cancel each other out. Also, auditors in the *NO REWARD* condition appropriately adjusted their *PS* based on available evidence, whereas auditors in the *REWARD* condition did not (illustrated in Table 1, Panel B and Figure 1).

---

<sup>13</sup> Inferences are identical if we exclude auditors who answered the manipulation check incorrectly.

Specifically, *PS* in the *NO REWARD* condition was higher for the severe, as opposed to minor *RF* (0.70 versus 0.35,  $p < 0.05$ ). By contrast, *PS* was approximately the same in the *REWARD* conditions, regardless of *RF* severity. As such, greater sensitivity to *RF* in the *NO REWARD* condition yields similar univariate means across *REWARD* conditions. We discuss this finding further when reporting tests for H2 examining the appropriateness of skepticism.

### 3.6.3. Tests of Hypotheses 1A and 1B – Skeptical Behavior

Table 2, Panel A reports the results of our tests of H1A and H1B. H1A predicts that a reward for costly skepticism increases auditors' perceived personal benefits and, in turn, decreases auditors' skepticism on a subsequent task. H1B predicts that a reward for costly skepticism increases auditors' motivation to engage in skepticism and, as such, increases auditors' subsequent skepticism. In tandem, H1A and H1B predict that *REWARD* has two opposite-signed indirect effects. To test these hypotheses, we use the following regressions:

$$BENEFIT = \delta_1 + \beta_1 REWARD + \varepsilon \quad (1)$$

$$MOTIVATION = \delta_2 + \beta_2 REWARD + \varepsilon \quad (2)$$

$$PS (1 \text{ or } 0) = \delta_3 + \beta_3 REWARD + \beta_4 BENEFIT + \beta_5 MOTIVATION + \varepsilon \quad (3)$$

We use the Preacher and Hayes (2008) bootstrapping approach to test the indirect effects predicted by H1A and H1B. We use 5,000 bootstrap re-samples with replacement to estimate 90% confidence intervals for each indirect effect, with significance indicated by intervals that do not include zero. The indirect effect predicted by H1A is the product  $\beta_1 REWARD * \beta_4 BENEFIT$ . We expect this product to be negative. The indirect effect predicted by H1B is the product of  $\beta_2 REWARD * \beta_5 MOTIVATION$ . We expect this product to be positive.<sup>14</sup>

---

<sup>14</sup> Our hypothesis test inferences are identical using 95% confidence intervals. However, 90% confidence intervals are often *ex ante* appropriate when using a dichotomous dependent variable that inherently inflates type 2 error (Demidenko [2007]). Also, note that equations (1) and (2) are OLS regressions and equation (3) is a logistic regression. The Preacher and Hayes [2008] procedure uses standardized coefficients to compute confidence intervals because the indirect effects involve multiplying coefficients from OLS with coefficients from logit.

As shown in Table 2 and Figure 2, the results support H1A. Auditors believe that a *REWARD* for costly skepticism increases their *BENEFIT* ( $p < 0.01$ ) (column i), and the effect of *BENEFIT* on *PS* is significantly negative ( $p < 0.05$ ) (column iii). In addition, the product of  $\beta_1 \text{REWARD} * \beta_4 \text{BENEFIT}$  is negative and significant (Lower CI = -2.08, Upper CI = -0.44). Auditors believe that a reward for costly skepticism provides a positional advantage over peers and are thus *less willing* to engage in subsequent skeptical actions. However, the results do not support H1B, as auditors do not assess their *MOTIVATION* as higher in response to a *REWARD* ( $p > 0.10$ ) as seen in Table 2 (column ii).<sup>15</sup> Figure 2 reports that the product of  $\beta_1 \text{REWARD} * \beta_5 \text{MOTIVATION}$  is insignificant (Lower CI = -0.09, Upper CI = 0.25).<sup>16</sup>

### 3.6.4. Estimating the Negative Effects of Rewards for Costly Skepticism

Note in Table 2 that *REWARD* has a marginally positive coefficient on *PS* after controlling for the mediators (column iii), which indicates a positive *direct effect* of *REWARD* on *PS* (Preacher and Hayes [2008]). The positive direct effect of *REWARD* indicates that rewards for costly skepticism have the *potential* to increase the application of skepticism, if one could remove the negative *BENEFIT* effect. On this note, to quantify the indirect effects on *PS*, Pearl ([2005]; [2010]) suggests a procedure to estimate “what could have been” if the negative effect of a mediator had not occurred. That is, it involves estimating what *PS* would have been in the *REWARD* condition using the mediator values from the *NO REWARD* condition. Untabulated results related to this procedure suggest *PS* would be 0.75 in the *REWARD* condition if the negative effect of *BENEFIT* would not occur, versus the 0.47 we observe in Table 1, Panel A. This difference of 0.28 is referred to as the “natural indirect effect.” Thus, the natural indirect

<sup>15</sup> Although the effect of *MOTIVATION* on *PS* is, as one would expect, significantly positive ( $p < 0.05$ ) (column iii), it is not via an indirect effect of *REWARD*.

<sup>16</sup> In equation (3), variance inflation factors for *BENEFIT*, *MOTIVATION*, and *REWARD* are less than or equal to 2.4. We re-perform individual regressions with only *BENEFIT* and *MOTIVATION* as predictors of *PS* and find that the inferences do not change, thus multicollinearity does not affect our findings.

effect of *REWARD* via *BENEFIT* in our study is a decrease in *PS* of approximately 0.28 in the *REWARD* condition. This illustrates that rewards *could* increase skepticism, if auditors' dysfunctional reactions to these evaluations could be undone.

#### *3.6.5. Auditors' Beliefs about Rewards for Costly Skepticism*

Our theory implicitly assumes that auditors may believe performance evaluations regress towards the mean (i.e., those receiving a reward (as opposed to no reward) may be less likely to expect their next evaluation outcome to be positive). The data supports our theory that auditors interpret the reward as a non-credible, better-than-expected outcome that is unlikely to be repeated. To test this possibility, we first measure (1) auditors' self-evaluations of their performance for the task, and (2) auditors' predictions of how the manager would evaluate their performance. We measure each on 11-point Likert scales anchored by -5 = "Not meeting Expectations," 0 = "Meeting Expectations," and +5 = "Exceeding Expectations." If self-evaluations are higher than predicted manager evaluations, then auditors expect a worse outcome than they deserve (i.e., a negative outcome).

Consistent with our assumption, receiving a reward for costly skepticism causes auditors to expect a negative evaluation for the subsequent task. Self-evaluations are significantly higher than predicted manager evaluations in the *REWARD* condition ( $p < 0.01$ , untabulated). By contrast, self-evaluations are lower than predicted manager evaluations in the *NO REWARD* condition ( $p = 0.07$ , untabulated), indicating expectations of a more positive outcome. In brief, receiving a reward on one task caused auditors to expect a negative outcome on the next task, while not receiving a reward caused auditors to expect a positive outcome on the next task.

#### *3.6.6. Tests of Hypothesis 2 – Appropriateness of Skeptical Behavior*

As shown in Panel A of Figure 3, H2 predicts that the indirect effects of *REWARD* via *BENEFIT* and *MOTIVATION* on *PS* will be conditional on *RF* severity. That is, H2 predicts “moderated mediation” in which the mediating effects are conditional on the evidence set available to the auditor. We test this hypothesis by estimating the indirect effects predicted by H1A and H1B (as described in the previous section) at each level of *RF* severity.<sup>17</sup> As shown in Panel B of Table 2, neither hypothesized indirect effect is conditional on *RF* severity in the manner that we predict. Specifically, the dysfunctional self-interest effect through *BENEFIT* is strongest when *RF* is severe (the only interval in Panel B that does not include zero), which is opposite of our prediction. This significantly negative effect reflects the fact that the greatest *PS* difference between the *NO REWARD* and *REWARD* conditions is in the severe *RF* condition (0.70 versus 0.44 as depicted in Table 1, Panel B and Figure 1). Thus, the *BENEFIT* effect constrains skepticism most clearly in the severe *RF* condition, which is very alarming for audit quality. Moreover, the indirect effect through *MOTIVATION* is insignificant at all levels of *RF*. Thus, H2 is not supported.

A plausible explanation for this finding is that auditors believe skeptical action would be more costly in the severe *RF* condition than in other conditions (e.g., a larger inconsistency requires more testing). To test this possibility, we measured auditors’ beliefs that additional procedures would cause a budget overrun on a 10-point Likert scale. Auditors assessed a higher likelihood that additional procedures would go over budget in the severe *RF* condition than in the moderate or minor *RF* conditions ( $p = 0.02$ , untabulated). This supports our logic that auditors in

---

<sup>17</sup> We use this approach for simplicity of presentation. An alternative test is to re-estimate equation 3 with an interaction term between *RF* and each of the mediators (i.e.,  $PS (1 \text{ or } 0) = \delta_4 + \beta_6 REWARD + \beta_4 MOTIVATION + \beta_7 BENEFIT + \beta_8 MOTIVATION*RF + \beta_9 BENEFIT*RF + \beta_{10} RF$ ). The predicted moderated mediation is determined by observing a significantly negative coefficient on the *BENEFIT\*RF* interaction term and/or a significantly positive coefficient on the *MOTIVATION\*RF* interaction term. We observe neither, and our inferences are identical using either approach.

the *REWARD* condition are strongly averse to risking poor objective performance (e.g., budget overruns), because auditors in the *NO REWARD* condition act skeptically in the severe *RF* condition even though they recognize the costs.

### 3.7 ADDITIONAL ANALYSES

#### 3.7.1. Recognition of the Red Flag and Informing the Audit Manager

While *PS* measures our participants' skeptical *actions*, we separately examine their skeptical *judgments*, which should influence *PS* (Nelson 2009). We find that auditor hesitation to act on the severe *RF* in the *REWARD* condition is not driven by insufficiently skeptical judgments. Recall that auditors in both *REWARD* conditions passed the manipulation check for *RF*, thus auditors noticed the negative NFM trend. Further, we measured *RECOGNITION* of the red flag as the degree of perceived inconsistency between reported sales and NFMs on a 10-point Likert scale with endpoints of 1 (Very Small) and 10 (Very Large). Table 3 reports the cell means of *RECOGNITION*.

Auditors in both the *REWARD* and *NO REWARD* conditions appear to have made appropriate skeptical judgments. That is, *RECOGNITION* increases from minor to severe *RF*. Nevertheless, when linking these judgments to action, we again observe that *REWARD* plays a dysfunctional role. As Nelson (2009) would suggest, *RECOGNITION* is positively correlated with *PS*, but only in the *NO REWARD* condition ( $\rho = +0.37$ ). We see no such link between judgment and action in the *REWARD* condition ( $\rho = -0.06$ ). Being aware of, but not appropriately considering, contradictory audit evidence is a behavior consistent with PCAOB inspection findings (PCAOB 2012).

Finally, we find that inappropriate responses could threaten audit quality by constraining auditors' willingness to communicate red flags to their manager. Before auditors made their *PS*

choice, they documented any issues that they would discuss with their manager. We coded a variable *INFORM MANAGER* that equals “1” if the auditor noted that they would discuss the NFM trend with their manager and “0” otherwise. Table 3 and Figure 1 report and illustrate cell means, respectively. In short, in the *NO REWARD* condition we find that *INFORM MANAGER* increases as the *RF* moves from minor to severe (0.31 versus 0.79,  $p < 0.01$ ). We do not observe such a positive relation when *REWARD* is present (0.44 versus 0.53,  $p = 0.62$ ).<sup>18</sup> The dysfunctional response to rewards of costly skepticism also decreases communication of red flags up the audit team hierarchy, further compromising audit quality (Lightle et al. 2017).

### 3.7.2. Auditors’ Personal History of Being Rewarded for Costly Skepticism

To this point, our findings have been discouraging. Avoiding skeptical actions likely reflects the belief that costly skepticism will not be rewarded, but natural differences in this belief can develop through experience. Auditors work on multiple engagements for a variety of supervisors in any given year, and supervisors vary in their evaluation biases and preferences (e.g., Tan and Jamal 2001). Auditors with a history of being rewarded for costly skepticism likely have stronger beliefs that costly skepticism *will be rewarded*, and thus may be more willing to engage in skepticism. That is, such a history may increase *MOTIVATION* for skepticism and constrain the *BENEFIT* effect.

Therefore, we test whether a history of being rewarded for costly skepticism moderates the indirect effects observed in H1A and H1B. We post-experimentally measured the variable *REWARD HISTORY* asking the auditors whether or not their managers would reward them for exercising professional skepticism when “You are over budget, the relationship with management is strained, and your skeptical behavior DID NOT identify a misstatement.”

---

<sup>18</sup> An untabulated mediation analysis confirms that the *REWARD* to *BENEFIT* to *PS* indirect effect (H1A) also decreases *INFORM MANAGER*.

Participants responded on a 10-point Likert scale with endpoints ranging from 1 (Would Not Reward Skepticism) to 10 (Would Definitely Reward Skepticism). To ease interpretation and the estimation of cell means, we create a dichotomous variable partitioning our sample on *REWARD HISTORY* in the top third versus bottom two thirds.<sup>19</sup>

As depicted in Figure 3, Panel B, we test whether *REWARD HISTORY* moderates the relations between *REWARD* and our two mediators *BENEFIT* and *MOTIVATION*. In simple terms, we are interested in whether the negative self-interest effect through *BENEFIT* decreases, and/or the positive effect through *MOTIVATION* increases, if we consider natural differences in *REWARD HISTORY*.

We estimate the following equations:

$$BENEFIT = \delta + \beta_6 REWARD + \beta_7 REWARD HISTORY + \beta_8 REWARD * REWARD HISTORY + \varepsilon \quad (4)$$

$$MOTIVATION = \delta + \beta_9 REWARD + \beta_{10} REWARD HISTORY + \beta_{11} REWARD * REWARD HISTORY + \varepsilon \quad (5)$$

$$PS (1 \text{ or } 0) = \delta + \beta_{12} REWARD + \beta_{13} MOTIVATION + \beta_{14} BENEFIT + \varepsilon \quad (6)$$

We provide descriptive statistics partitioned by *REWARD HISTORY* and *REWARD* in Table 4, Panel A. As shown in Panel B column (i), *REWARD HISTORY* does not moderate the indirect effect of *REWARD* on *PS* via *BENEFIT*. That is, there is no effect of *REWARD HISTORY* or the *REWARD \* REWARD HISTORY* interaction.<sup>20</sup>

The findings for motivational effects are encouraging. As shown in Panel B of Table 4, in column (ii) *REWARD HISTORY* is significantly positively associated with *MOTIVATION* ( $p < 0.01$ ). There is also a marginally significant *REWARD \* REWARD HISTORY* interaction term ( $p < 0.10$ ). The interaction is consistent with the descriptive means in Table 4 Panel A. Specifically,

<sup>19</sup> Our inferences do not change if we use the continuous measure, but estimating cell means and interpreting marginal effects is more straightforward when using a dichotomous measure.

<sup>20</sup> Although a history of being rewarded for costly skepticism by a supervisor may prompt an auditor to exercise skepticism on *that supervisor's* engagement, such experience may not do so with *other* supervisors since auditors have strong priors that, in general, rewarding costly skepticism is relatively uncommon.



in the *NO REWARD* condition, auditors who have a positive *REWARD HISTORY* have significantly higher *MOTIVATION* (8.53 vs. 6.83). By contrast, when *REWARD* is present, *MOTIVATION* is relatively high regardless of *REWARD HISTORY*. This is consistent with *REWARD HISTORY* helping auditors maintain the motivation to be skeptical, despite *not* being rewarded for costly skepticism on an earlier task.

These results further highlight the benefits of a shift in evaluation culture that reliably rewards costly, yet appropriate skepticism. Auditors whose supervisors provide credible rewards for costly skepticism are motivated to be skeptical, even when they receive a negative one-off evaluation of costly skepticism. In an untabulated analysis, we include *REWARD HISTORY* in a regression predicting *PS* and find that it is positively associated with *PS*, controlling for our other predictors.<sup>21</sup> Notably, we also measure auditors' history of rewards for *general* skepticism (where the outcome of the skepticism was not noted in the measure) and find no relation with *PS* in our experimental task. This suggests higher skepticism may not stem from a history of general rewards for skepticism, but rather from the experience of being rewarded specifically for *costly* skepticism. If a positive *REWARD HISTORY* with costly skepticism is important for encouraging skepticism, then a natural question emerges: what leads a supervisor to reliably reward costly skepticism? We address this question in our follow-up survey.

#### *4. Auditor Survey – Who Rewards Costly Skepticism?*

In our experiment, we find that auditors who have more experience being rewarded for costly skepticism are more motivated to engage in skepticism, even when they do not receive a reward in the experiment. Thus, we conduct a case-based survey to explore the attributes of audit

---

<sup>21</sup> In untabulated analyses, we also measure and control for other determinants suggested by Nelson (2009) including trait skepticism (Hurt 2010) and various dimensions of knowledge such as industry experience, experience with analytical procedures, and experience with NFMs. Our inferences do not change when controlling for these variables.

supervisors (e.g., traits, experiences) who are likely to reliably reward costly skepticism. We use a case-based survey due to the potential for social desirability bias in a standard survey (i.e., supervisors may not respond honestly to a standard survey question asking whether or not they reward costly skepticism). Further, because traits and experiences are relatively stable attributes across situations (Kassin [2003]), evidence that a participant rewards costly skepticism in our study is a valid signal that the participant would do so in other settings. Identifying the factors that are associated with supervisors who provide positive evaluations of costly skepticism can provide insights into how these types of supervisors can be recruited and/or developed.

#### 4.1 PARTICIPANTS AND PROCEDURES

Participants in our survey were 127 practicing auditors from two Big Four and two non-Big Four firms. We collected responses from two of the firms at firm-sponsored training sessions and from the other two firms via an online survey hosted by Qualtrics. There are no significant effects for firm or data collection medium. Participants' mean auditing experience was 6.3 years, with 95% indicating experience evaluating the performance of subordinates.

All participants received the exact same case-based content in their research materials. The materials are adapted from Brazel et al. [2016] and asked auditors to *evaluate* the performance of a staff member under their supervision at a hypothetical audit client in the manufacturing industry. The materials then described the staff member's performance on a substantive analytical procedure related to a division's sales. Similar to our auditor experiment, the staff auditor noted a red flag between trends in reported sales and trends in non-financial measures (NFM), whereas financial measures were consistent with sales.

Participants were informed that the staff auditor incorporated NFMs into the analytical procedure for the current year, noted an inconsistency between sales growth and related NFMs,

and chose to investigate the red flag. The identification and investigation of the red flag caused the staff auditor to go over budget and strain relations with management. The staff auditor ultimately determined that the inconsistency was the result of the division outsourcing some operations overseas. Additional testing related to the overseas operations led to a conclusion that there was no misstatement in the sales account. Participants were then asked to evaluate the staff auditor, who had engaged in appropriate, yet costly skepticism.<sup>22</sup>

## 4.2 MEASURES

The primary dependent measure is participants' evaluation of the staff member (*EVAL*). We measure *EVAL* with the question "Based on the information presented on the prior pages, how would you evaluate [the staff auditor's] overall performance?" Participants responded on an 11-point Likert scale with endpoints ranging from -5 (Below Expectations) to +5 (Above Expectations). We follow the Nelson [2009] model to identify and categorize measures that could influence skeptical behavior/the evaluation of skepticism, which we group into incentives, traits, and knowledge.<sup>23</sup> The Appendix provides details on the measurement of these variables. For incentives, we measure beliefs about how the staff member's actions will affect the evaluator's own performance evaluation by their audit partner with *AFFECT OWN EVAL*. We also measured *GENERAL PS REWARDS*, how participants' own managers in the past had rewarded *general* skepticism. Finally, similar to our experiment, we measured *REWARD HISTORY* for costly skepticism. For traits, we measured the *TRAIT SKEPTICISM* of participants

---

<sup>22</sup> Our survey participants' mean rating for the appropriateness of the subordinate's skepticism was 1.05 on a scale ranging from -5 (Was NOT SUFFICIENTLY skeptical) to 0 (Was APPROPRIATELY Skeptical) to +5 (was OVERY skeptical).

<sup>23</sup> Nelson (2009) also includes "evidential input" as a determinant, but our research materials hold constant the evidential input related to the analytical procedure.

using the Hurtt [2010] scale. We also measured participants' beliefs about their own general abilities by measuring *MY RANK* relative to others at the same experience level.<sup>24</sup>

Nelson [2009] describes how relevant knowledge is developed through experiences and training. Accordingly, we measured multiple dimensions of experience and training that could influence knowledge related to our setting. We measured general audit *EXPERIENCE*, *NFM* experience, and manufacturing *INDUSTRY* experience. We also measured participants' *FRAUD TRAINING* and their perceptions of the misstatement base rate (*MM BASERATE*), as perceptions of base rates likely influence skepticism (e.g., Bonner et al. [1997]). Finally, evaluators' own experiences with consultative and supportive supervisors can affect how they evaluate skepticism (Glover and Prawitt [2014]; Nelson et al. [2016]). We therefore include the measure *CONSULTATIVE* to capture the extent to which participants consulted with their own supervisors while exercising skepticism.

#### 4.3 RESULTS

Table 5 presents the results of multivariate analyses in which we run separate OLS regressions for each group of determinants, as specified in Nelson [2009], as well as a full model of determinants on *EVAL*. In column (i), we find *AFFECT OWN EVAL* is positively associated with rewarding costly skepticism. Thus, our evidence highlights the importance of “tone at the top,” as audit partners who endorse costly skepticism on their engagements are likely to develop supervisors who reward their staff for the behavior. In contrast to our experimental results for

---

<sup>24</sup> Although this measure has some similarity with the BENEFIT measure used in the experiment as they both evaluate perceptions of ranking, this measure differs on two important dimensions: (1) it captures participants' perceptions of their overall general ability rather than their ability in regards to performance in a single evaluation setting, and (2) it is *not conditioned* on anything in the research materials.

subordinates, we do not find an effect of *evaluators'* own experiences being rewarded for costly skepticism (i.e., no effect for *REWARD HISTORY*).<sup>25</sup>

In column (ii), note that we find no effect for *TRAIT SKEPTICISM*. However, auditors who have higher general ability tend to reward costly skepticism, as evidenced by the significant and positive coefficient for *MY RANK*. If auditors ranked higher in their class are more influential at their firms, then it is possible that they could be effective in sharing/training “best practices” such as reliably rewarding costly skepticism.<sup>26</sup>

In column (iii), we find *CONSULTATIVE* has a positive and significant sign, consistent with the idea that more positive evaluations of costly skepticism stem from auditors who experience more supportive and open environments in which auditors feel comfortable raising issues. Finally, we find two knowledge effects in which total *EXPERIENCE* and *NFM* are positively associated with evaluations of costly skepticism.

In sum, our survey findings are consistent with the intuition of our experimental findings. Beliefs that superiors will reward or support costly skepticism are associated with (1) increased motivation for skepticism among subordinates and (2) more positive evaluations of costly skepticism by supervisors. Similarly, among evaluators, experiencing more consultative supervisors is associated with higher evaluations of skepticism. Just as subordinates were more willing to act skeptically when they had credible incentives for costly skepticism, evaluators appear more willing to reward costly skepticism when they have experienced audit teams that valued a questioning mind/consultation within the team. Overall, there appears to be a reciprocal,

---

<sup>25</sup> Reasoning *ex post*, this difference may be attributable to the survey asking how their *audit partner* would react to costly skepticism and the experiment asking how their *audit manager* would react to costly skepticism. Agoglia et al. (2015) find that managers are more sensitive to budget overages than partners are, thus the incentive intensity captured by this particular measure may be greater in the experiment than in the survey.

<sup>26</sup> This finding also raises the encouraging possibility that when auditors have relatively stable beliefs about their high ranking, they may be more willing to promote skepticism. By contrast, the experimental finding that higher rankings lead to lower skepticism would still apply when auditors believe their higher ranking directly results from an unexpected benefit.

“pay it forward” mentality cultivated by strong mentoring environments that reliably reward costly skepticism.

## 5. *CONCLUSIONS*

We examine auditors’ incentives to exercise professional skepticism. We begin with the widely-held premise that there is insufficient skepticism in practice and recent evidence that auditor incentive systems do not adequately reward costly skepticism (appropriate skeptical behavior that generates costs, but identifies no misstatement). In our experiment, auditors rewarded for costly skepticism interpret a positive evaluation as a benefit that they wish to preserve and, in turn, are not apt to exercise skepticism in a subsequent task. On the positive side, those who have experienced consistent rewards for costly skepticism are more motivated to exercise skepticism, even after they receive a negative evaluation for costly skepticism. A supplemental survey finds that supervisors are most likely to reward costly skepticism when they have experienced consultative supervisors who are themselves supportive of costly skepticism. Consequently, audit firms are likely to benefit from developing cultures in which appropriate skepticism is encouraged, regardless of outcome.

However, our findings also suggest that in order to increase skepticism in the field, incentives for exercising skeptical must be credible for the individual auditor. Our results indicate that the current state of performance evaluation systems in audit firms may create dysfunctional incentives in which audit subordinates may not increase skepticism even when prior skepticism has been rewarded. The idea that auditors are strategic is consistent with the review process literature showing that auditors strategically manage impressions with their reviewers (e.g., Rich et al. [1997]). It is also consistent with the stream of research on the unintended effects of bias in auditor supervision and review (e.g., Tan and Jamal [2001]; Wilks

[2002]; Peecher et al. [2010]). While reviewers have developed mechanisms to cope with strategic behavior (e.g., Tan and Trotman [2003]), there is not an immediately obvious response to the dysfunctional effect we observe. Given evidence that audit supervisors have poor insights into subordinates' capabilities (Kennedy and Peecher [1997]; Messier et al. [2008]; Peecher et al. [2010]), it is unlikely that supervisors have good insight into subordinates' motivations either. Further, we find that rewards for costly skepticism decrease the likelihood that auditors would communicate issues to their supervisors, which limits supervisors' abilities to respond. Future research could examine the conditions under which audit supervisors anticipate this strategic, non-skeptical behavior by their subordinates and, if they do, the mechanisms they employ to cope with such behavior.

In addition, when skeptical behavior yields a misstatement the outcome is documented in the audit adjustment file. When skepticism does not detect a misstatement, it is uncertain whether and how this exercise of skepticism is visible to others. Better documentation of costly skepticism could reduce the negative effects we observe. Future research can examine ways in which costly skepticism can be documented (e.g., in the budget file) such that acts of costly skepticism can be effectively conveyed in audit the documentation.

## REFERENCES

- AGOGLIA, C., R. HATFIELD, AND T. LAMBERT. "Audit team time reporting: An agency theory perspective." *Accounting, Organizations and Society* 44 (2015): 1 – 14.
- BAIMAN, S. "Agency research in managerial accounting: A second look." *Accounting, Organizations and Society* 15(4) (1990): 341-371.
- BERGER, J., C. HARBRING, AND D. SLIWKA. "Performance appraisals and the impact of forced distribution: An experimental investigation." *Management Science* 59 (1) (2013): 54 – 68.
- BLANKEN, I., N. VAN DEN VEN, AND M. ZEELLENBERG. "A meta-analytic review of moral licensing." *Personality and Social Psychology Bulletin* 41(5) (2015): 540 – 558.
- BOL, J. "Subjectivity in compensation contracting." *Journal of Accounting Literature* 27 (2008): 1-32.
- BOL, J. C., AND S. SMITH. "Spillover effects in subjective performance evaluation: Bias and the asymmetric influence of controllability." *The Accounting Review* 86(4) (2011): 1213-1230.
- BOL, J., AND J. LEIBY. "Subjectivity in professionals' incentive systems: Differences between promotion- and performance-based assessments." *Contemporary Accounting Research* (forthcoming) (2018).
- BONNER, S.E., R. LIBBY, AND M. NELSON. "Audit category knowledge as a precondition to learning from experience." *Accounting, Organizations and Society* 22 (5) (1997): 387–410.
- BONNER, S.E., AND G. SPRINKLE. "The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research." *Accounting, Organizations and Society* 27 (2002): 303 – 345.
- BRAZEL, J. F., K. L. JONES, AND M. F. ZIMBELMAN. "Using nonfinancial measures to assess fraud risk." *Journal of Accounting Research* 47 (December) (2009): 1135–1166.
- BRAZEL, J.F., K. L. JONES, AND D. PRAWITT. "Auditors' reactions to inconsistencies between financial and nonfinancial measures: The interactive effects of fraud risk assessment and a decision prompt." *Behavioral Research in Accounting* 26 (1) (2014): 131-156.
- BRAZEL, J.F., S. B. JACKSON, T. J. SCHAEFER, AND B. W. STEWART. "The outcome effect and professional skepticism." *The Accounting Review* 91 (6) (2016): 1577 – 1599.
- BRAZEL, J.F., C. GIMBAR, E. MAKSYMOW, AND T. J. SCHAEFER. "The outcome effect and professional skepticism: A replication and an attempt at mitigation. Working Paper, North Carolina State University.
- CAPPELLI, P., AND A. TAVIS. "The performance management revolution." *Harvard Business Review* 94(10) (2016): 58 – 67.
- CARDINAELS, E., AND H. YIN. "Think twice before going for incentives: Social norms and the principal's decision on compensation contracts." *Journal of Accounting Research* 53(5) (2015): 985 – 1015.
- CHRISTENSEN, B. AND N. NEWTON. "How does audit team staffing affect audit outcomes? Archival evidence from U,S, audits." Working paper, University of Missouri (2017).
- CICHELO, M. S., C. E. FEE, C. J. HADLOCK, AND R. SONTI. "Promotions, turnover, and performance evaluation: Evidence from the careers of division managers." *The Accounting Review* 84 (4) (2009): 1119–1143.



- DEMIDENKO, E. "Sample size determination for logistic regression revisited." *Statistics in Medicine* 26 (2007): 3385 – 3397.
- DOPUCH, N., M. GUPTA, D. SIMUNIC, AND M. STEIN. "Production efficiency and the pricing of audit services." *Contemporary Accounting Research* 20 (1) (2003): 47–77.
- DURNEY, M., R. ELDER, AND S. GLOVER. "Field data on accounting error rates and audit sampling." *Auditing: A Journal of Practice & Theory* 33(2) (2014): 79 – 110.
- EARLEY, C. E. "Knowledge acquisition in auditing: Training novice auditors to recognize cue relationships in real estate valuation." *The Accounting Review* 76 (1) (2001): 81–97.
- EY. How big data and analytics are transforming the audit:  
<http://www.ey.com/gl/en/services/assurance/ey-reporting-issue-9-how-big-data-and-analytics-are-transforming-the-audit#item1> (2017)
- FEINTZEIG, R. "Your manager wants to see you. Again: Performance reviews at some companies now occur multiple times a year-even as often as every other week." *Wall Street Journal* (April 22) (2017): B7.
- GIBBS, M., K. MERCHANT, W. VAN DER STEDE, AND M. VARGUS. "Determinants and effects of subjectivity in incentives." *The Accounting Review* 79 (2004): 409-436.
- GLOVER, S. M., AND D. F. PRAWITT. "Enhancing auditor professional skepticism: The professional skepticism continuum." *Current Issues in Auditing*, 8 (2) (2014): 1-10.
- GRIFFITH, E., J. HAMMERSLEY, AND K. KADOUS. "Audits of complex estimates as verification of management numbers: How institutional pressures shape practice." *Contemporary Accounting Research* 32 (2015): 833 – 863.
- HACKENBRACK, K., AND C. HOGAN. "Client retention and engagement-level pricing." *Auditing: A Journal of Practice and Theory* 24 (1) (2005): 7 – 20.
- HOLMSTROM, B. "Contractual models of the labor market." *The American Economic Review* 71 (2) (1981): 308 – 313.
- HOUSTON, R. W. "The Effects of Fee Pressure and Client Risk on Audit Seniors' Time Budget Decisions." *Auditing: A Journal of Practice and Theory* 18 (1999): 70-86.
- HURTT, R. K. "Professional skepticism: An audit specific model and measurement scale." *Auditing: A Journal of Practice and Theory* 29 (1) (2010):149-171.
- INTERNATIONAL AUDITING AND ASSURANCE STANDARDS BOARD (IAASB). *Objective and general principles governing an audit of financial statements*. ISA 200. New York, NY (2004): IFAC.
- INTERNATIONAL AUDITING AND ASSURANCE STANDARDS BOARD (IAASB). "Quality control for firms that perform audits and reviews of financial statements, and other assurance and related services engagements. " ISQC 1. New York, NY (2009): IFAC.
- INTERNATIONAL AUDITING AND ASSURANCE STANDARDS BOARD (IAASB). *Professional Skepticism in an Audit of Financial Statements*. New York, NY (2012): IFAC.
- INTERNATIONAL AUDITING AND ASSURANCE STANDARDS BOARD (IAASB). *Invitation to comment: Enhancing audit quality in the public interest: A focus on professional skepticism, quality control and group audits*. New York, NY (2015): IFAC.
- INTERNATIONAL AUDITING AND ASSURANCE STANDARDS BOARD (IAASB). *Work Plan for 2017-2018: Enhancing Audit Quality*. New York, NY (2017): IFAC.

- KADOUS, K., J. KENNEDY, AND M. PEECHER. "The Effect of Quality Assessment and Directional Goal Commitment on Auditors' Acceptance of Client-Preferred Accounting Methods." *The Accounting Review* 78(3) (2003): 759 – 778.
- KAHNEMAN, D., AND A. TVERSKY. "On the psychology of prediction." *Psychological Review* 80(4) (1973): 237 – 251.
- KAHNEMAN, D., AND A. TVERSKY. "Prospect theory: An analysis of decision under risk." *Econometrica* 47 (2) (1979): 263-291.
- KAPLAN, S., AND P. RECKERS. "An examination of information search during initial audit planning." *Accounting, Organizations and Society* 14 (1989): 337 – 345.
- KAPLAN, S., C. MOECKEL, AND J. WILLIAMS. "Auditors' hypothesis plausibility assessments in an analytical review setting." *Accounting, Organizations and Society* 11 (1992): 50 – 65.
- KASSIN, S. *Psychology* 4<sup>th</sup> Edition (2003) Upper Saddle River, NJ: Prentice Hall.
- KENNEDY, J. "Debiasing audit judgment with accountability: A framework and experimental results." *Journal of Accounting Research* Autumn (1993): 231 – 245.
- KENNEDY, J., AND M.E. PEECHER. "Judging auditors' technical knowledge." *Journal of Accounting Research* 35(2) (1997): 279 – 293.
- KPMG, LLP (KPMG). *Transparency Report 2016*. KPMG Professional Services (2016).
- KPMG. Audit data & analytics: <https://home.kpmg.com/xx/en/home/services/audit/audit-data-analytics.html>. (2017).
- LIBBY, R., AND J. LUFT. "Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment." *Accounting, Organizations and Society* 18(5) (1993): 425 – 450.
- LIGHTLE, S., J. CATELLANO, AND B. BAKER. "Why audit teams need the confidence to speak up." *Journal of Accountancy* (January) (2017): 46-51.
- MACLEOD, W. "Optimal contracting with subjective evaluation." *The American Economic Review* 93 (March) (2003): 216-240.
- MAYHEW, B. "Auditor reputation building." *Journal of Accounting Research* 39 (3) (2001): 599 – 617.
- MESSIER, W., V. OWHOSO, AND C. RAKOVSKI. "Can audit partners predict subordinates' ability to detect errors?" *Journal of Accounting Research* 46(5) (2008): 1241 – 1264.
- NELSON, M.W. "A model and literature review of professional skepticism in auditing." *Auditing: A Journal of Practice & Theory* 28 (2) (2009): 1-34.
- NELSON, M.W., C.A. PROELL, AND A. RANDEL. "Team-oriented leadership and auditors' willingness to raise audit issues." *The Accounting Review* 91(6) (2016): 1781 – 1805.
- PEARL, J. "Direct and indirect effects." *Proceedings of the American Statistical Association Joint Statistical Meeting*. R-273 (2005): 1572 – 1581.
- PEARL, J. An introduction to causal inference." *International Journal of Biostatistics* 6(2) (2010): 1 – 62.
- PEECHER, M.E., M. D. PIERCEY, J. S. RICH, AND R. M. TUBBS. "The effects of a supervisor's active intervention in subordinates' judgments, directional goals, and perceived technical knowledge advantage on audit team judgments." *The Accounting Review* 85 (5) (2010): 1763–1786.
- PEECHER, M. E., I. SOLOMON, AND K. T. TROTMAN. "An Accountability Framework for Financial Statement Auditors and Related Research Questions." *Accounting, Organizations and Society* 38 (2013): 596-620.

- PREACHER, K., AND A. HAYES. "Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models." *Behavioral Research Methods* 40 (3) (2008): 879-891.
- PRENDERGAST, C. "The provision of incentives in firms." *Journal of Economic Literature* 37 (1) (1999): 7-63.
- PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Maintaining and Applying Professional Skepticism in Audits*. Staff Audit Practice Alert No. 10 (SAPA 10). Washington, DC (2012): PCAOB.
- PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Due Professional Care in the Performance of Work: Auditing standard 1015*. Washington, DC (2015): PCAOB.
- PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Consideration of fraud in a financial statement audit: Auditing standard 2401*. Washington, DC (2016): PCAOB.
- PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Standard-Setting Update: Office of the Chief Auditor*. Washington, DC (2017): PCAOB.
- REITMAN, A. "Want to jettison the annual performance review? Here's how some organizations are making the change." *Journal of Accountancy* December (2017): 18-22.
- RICH, J.S., I. SOLOMON, AND K.T. TROTMAN. "The audit review process: A characterization from the persuasion perspective." *Accounting, Organizations and Society* 22 (5) (1997): 481 – 505.
- SHAUB, M. K., AND J. E. LAWRENCE. "Differences in Auditors' Professional Skepticism Across Career Levels in the Firm." *Advances in Accounting Behavioral Research* 2 (1999): 61-83.
- SOLOMON, I., M. SHIELDS, AND R. WHITTINGTON. "What do industry specialist auditors know?" *Journal of Accounting Research* 37 (1999): 191-208.
- TAN, H. T., AND K. JAMAL. "Do auditors objectively evaluate their subordinates' work?" *The Accounting Review* 76 (January) (2001): 99-110.
- TAN, H.T., AND J. YIP-OW. "Are reviewers' judgments influenced by memo structure and conclusions documented in audit workpapers?" *Contemporary Accounting Research* 18(4) (2001): 663 – 678.
- TAN, H. T., AND K.T. TROTMAN. "Reviewers' responses to anticipated stylization by preparers of audit workpapers." *The Accounting Review* 78 (2) (2003): 581 – 604.
- TAN, H. T., AND K. JAMAL. "Managing perceptions of technical competence: How well do auditors know how others view them?" *Contemporary Accounting Research* 23 (3) (2006): 761 – 787.
- TROMPETER, G., AND A. WRIGHT. "The world has changed—Have analytical procedure practices?" *Contemporary Accounting Research* 27 (2) (2010): 669-700.
- VAN SCOTTER, J., S. MOTOWILDO, AND T. CROSS. "Effects of task performance and contextual performance on systemic rewards." *Journal of Applied Psychology* 85(4) (2000): 526 – 535.
- WALLER, W., AND M. ZIMBELMAN. "A cognitive footprint in archival data: Generalizing the dilution effect from laboratory to field settings." *Organizational Behavior and Human Decision Processes* 91(2) (2003): 254 – 268.
- WERNER, J. "Dimensions that make a difference: Examining the impact of in-role and extrarole behaviors on supervisory ratings". *Journal of Applied Psychology* 79(1) (1994): 98-107.

WITT, L., AND G. FERRIS. "Social skill as moderator of the conscientiousness-performance: Convergent results across four studies." *Journal of Applied Psychology* 88(5) (2003): 809-820.

WILKS, T. J. "Predecisional distortion of evidence as a consequence of real-time audit review." *The Accounting Review* 77 (1) (2002): 51-71.

## APPENDIX – VARIABLE DESCRIPTIONS

| Variable                            | Description  |
|-------------------------------------|--|
| <b>EXPERIMENT</b>                   |  |
| <i>REWARD</i>                       | Manipulated between participants as an interim evaluation of “Met Expectations” (No Reward) or “Exceeded Expectations” (Reward).   |
| <i>RED FLAG (RF)</i>                | Manipulated at three levels between participants as MINOR, MODERATE, and SEVERE. In all conditions, sales growth was constant at positive 9%. In the MINOR condition, NFM growth was -1%. In the MODERATE and SEVERE conditions, NFM growth was -21% and -41%, respectively.   |
| <i>BENEFIT</i>                      | Self-assessments of how participants would rank in their class (compared to other seniors in their office) from receiving the evaluation in <i>REWARD</i> , collected on a scale ranging from 1 (Bottom of My Class) to 10 (Top of My Class).  |
| <i>MOTIVATION</i>                   | Self-assessments of how motivated participants were to exercise professional skepticism during the task. Participants responded on a scale ranging from 1 (Not at All Motivated) to 10 (Extremely Motivated).  |
| <i>PROFESSIONAL SKEPTICISM (PS)</i> | Coded as “1” if participants concluded that more work would be required in the task and “0” if participants concluded the balance appeared reasonable.   |
| <i>INFORM MANAGER</i>               | Coded as “1” if participants explicitly wrote down the NFM trend in the list of items they would communicate to their manager and “0” otherwise.   |
| <i>RECOGNITION</i>                  | Assessments of the extent to which the NFM trend is inconsistent with the growth in recorded sales, collected on a scale ranging from 1 (Very Small) to 10 (Very Large).   |
| <i>REWARD HISTORY</i>               | Self-assessment of likelihood that managers in the participant’s own personal experience would reward costly skepticism, collected on a scale ranging from 1 (Would Not Reward Skepticism) to 10 (Would Definitely Reward Skepticism). For analyses, we partition into HIGH (top third) and LOW (bottom two thirds). |
| <i>GENERAL SKEPTICISM</i>           | Self-assessment of likelihood that managers in participants’ own personal experience would reward skepticism, collected on a scale ranging from 1 (Would not reward skepticism) to 10 (Would definitely reward skepticism).  |
| <b>SURVEY</b>                       |  |
| <i>VAL</i>                          | Assessments of the staff member’s performance, collected on a scale ranging from -5 (Below Expectations) to 0 (Met Expectations) to +5 (Above Expectations).   |
| <i>AFFECT OWN VAL</i>               | Assessment of how the partner will evaluate the participant’s own performance collected on a scale ranging from -5 (Below Expectations) to 0 (Met Expectations) to +5 (Above Expectations).  |
| <i>GENERAL PS REWARDS</i>           | Self-reported experience being rewarded by managers for any skepticism, collected on a scale ranging from 1 (Would Not Reward Skepticism) to 10 (Would Definitely Reward Skepticism).  |
| <i>REWARD HISTORY</i>               | Self-assessment of likelihood that managers in the participant’s own personal experience would reward costly skepticism, collected on a scale ranging from 1 (Would Not Reward Skepticism) to 10 (Would Definitely Reward Skepticism). For analyses, we partition into HIGH (top third) and LOW (bottom two thirds). |
| <i>TRAIT SKEPTICISM</i>             | Trait skepticism measured via the Hurtt (2010) scale.  |
| <i>MY RANK</i>                      | Self-assessments of how participants currently rank in their class, collected on a scale ranging from 1 (Bottom of My Class) to 10 (Top of My Class).  |
| <i>CONSULTATIVE</i>                 | Self-reported extent to which participants consult their supervisors while exercising skepticism, collected on a scale ranging from 0 (Never) to 10 (Always).  |
| <i>TOTAL EXPERIENCE</i>             | Total self-reported audit experience in years & months (converted to decimals).  |
| <i>INDUSTRY</i>                     | The participant’s percentage of chargeable hours in manufacturing during the past three years (the industry in our case).  |
| <i>NFM</i>                          | Percentage of substantive analytical procedures conducted by the participant that have involved NFMs.  |
| <i>FRAUD TRAINING</i>               | The participant’s number of hours of training on financial statement fraud.  |
| <i>MM BASERATE</i>                  | The participant’s estimate of the percentage of pre-audit financial statements that contain a material misstatement.   |

FIGURE 1 – EXPERIMENT: CELL MEANS FOR SKEPTICISM & INFORM MANAGER

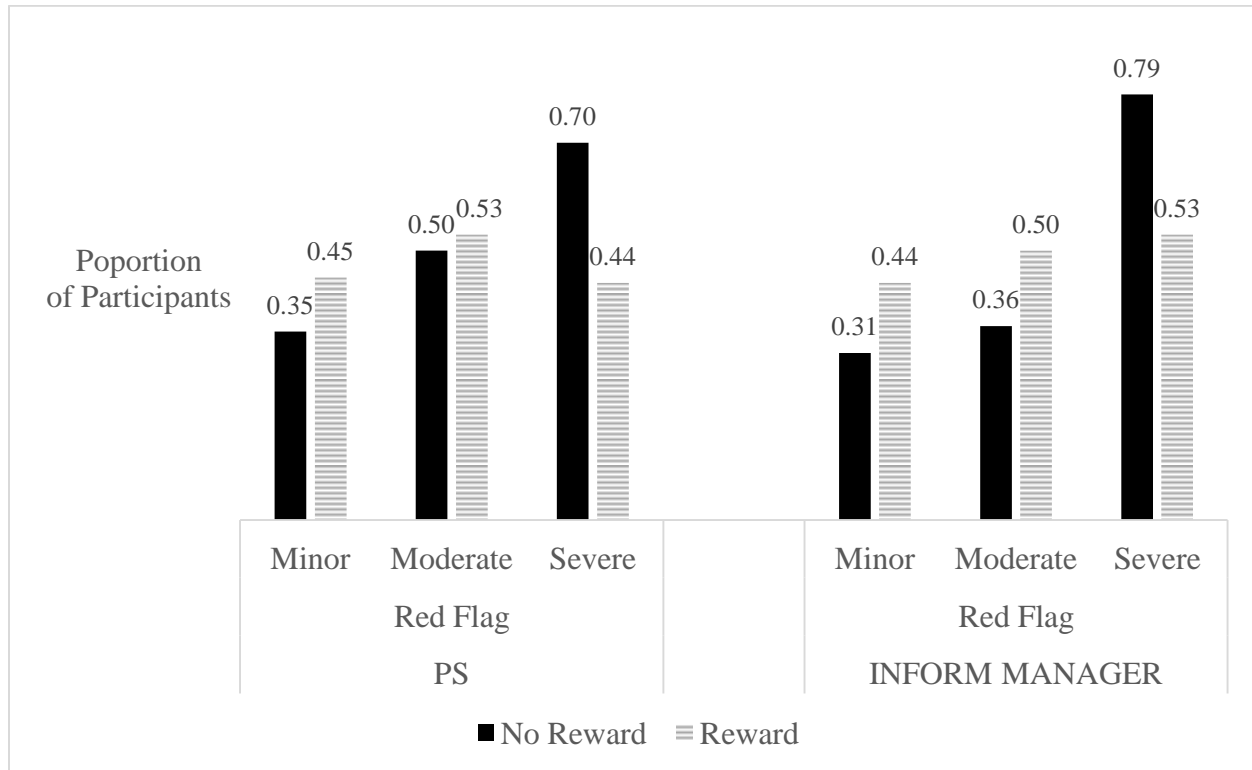


Figure 1 illustrates the cell means for *PROFESSIONAL SKEPTICISM (PS)* and *INFORM MANAGER* for each *RED FLAG* condition in each of the two *REWARD* conditions. *PROFESSIONAL SKEPTICISM (PS)* is coded as “1” if participants concluded that more work would be required in the task and “0” if participants concluded the balance appeared reasonable. *INFORM MANAGER* is coded as “1” if participants explicitly wrote down the NFM trend in the list of items they would communicate to their manager and “0” otherwise. *RED FLAG* was manipulated at three levels between participants as MINOR, MODERATE, and SEVERE. In all conditions, sales growth was constant at positive 9%. In the MINOR condition, NFM growth was -1%. In the MODERATE and SEVERE conditions, NFM growth was -21% and -41%, respectively. *REWARD* is manipulated between participants as an interim evaluation of “Met Expectations” (No Reward) or “Exceeded Expectations” (Reward).

FIGURE 2 – INDIRECT EFFECTS HYPOTHESIS TESTS

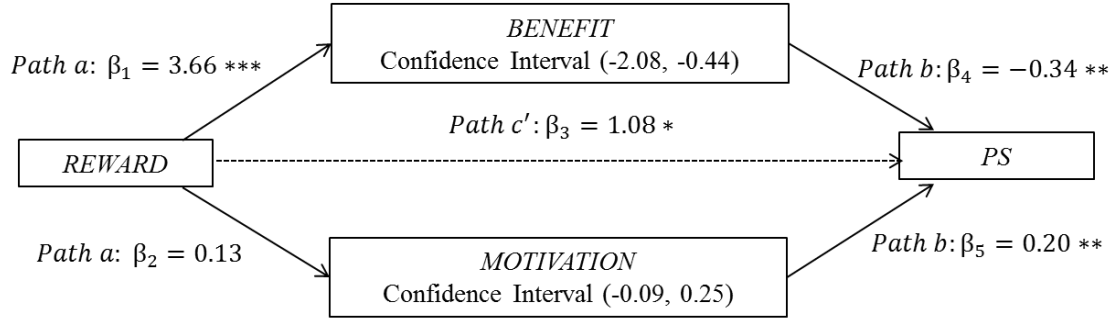


Figure 2 depicts the coefficients of the indirect effect of *REWARD* on *PROFESSIONAL SKEPTICISM (PS)* via *MOTIVATION* and *BENEFIT*. *REWARD* is manipulated between participants as an interim evaluation of “Met Expectations” (No Reward) or “Exceeded Expectations” (Reward). *BENEFIT* is a self-assessment of how participants would rank in their class (compared to other seniors in their office) from receiving the evaluation in *REWARD*, collected on a scale ranging from 1 (Bottom of My Class) to 10 (Top of My Class). *MOTIVATION* is a self-assessment of how motivated participants were to exercise professional skepticism during the task. Participants responded on a scale ranging from 1 (Not at All Motivated) to 10 (Extremely Motivated). *PROFESSIONAL SKEPTICISM (PS)* is coded as “1” if participants concluded that more work would be required in the task and “0” if participants concluded the balance appeared reasonable.

The coefficients are standardized to allow for computation of the indirect effects and are estimated using:

$$BENEFIT = \delta_1 + \beta_1 REWARD + \varepsilon \quad (1)$$

$$MOTIVATION = \delta_2 + \beta_2 REWARD + \varepsilon \quad (2)$$

$$PS (1 \text{ or } 0) = \delta_3 + \beta_3 REWARD + \beta_4 BENEFIT + \beta_5 MOTIVATION + \varepsilon \quad (3)$$

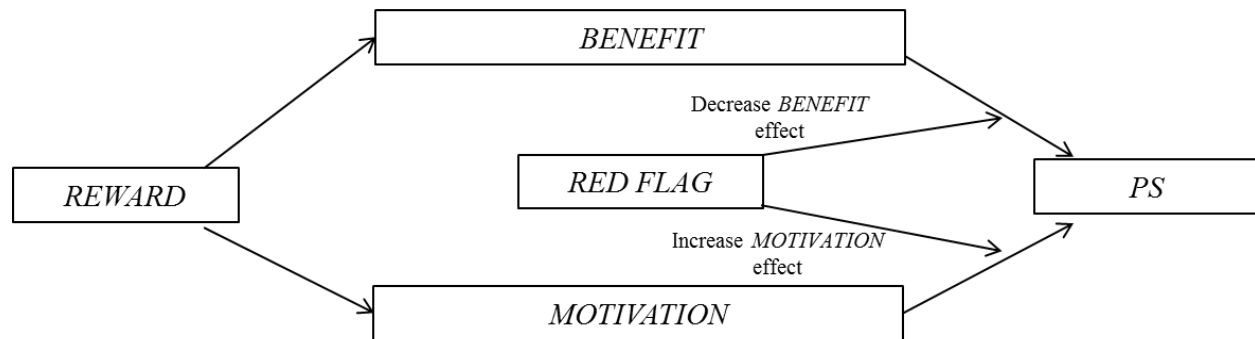
Significance of coefficients is indicated with \*\*\* for  $p < 0.01$ , \*\* for  $p < 0.05$ , and \* for  $p < 0.10$ .

The indirect effects of *REWARD* on *PS* are the products of  $\beta_1 REWARD * \beta_4 BENEFIT$  for the *BENEFIT* mediator and  $\beta_2 REWARD * \beta_5 MOTIVATION$  for the *MOTIVATION* mediator. Confidence intervals are bias-corrected intervals for the estimate of the indirect effect, which are estimated using 5,000 bootstrapped re-samples of the data with replacement. Significance of the indirect effect is indicated if the intervals do not include zero.

$\beta_3$  is from the path c' and is the *direct effect* of *REWARD* on *PS* controlling for the effects of *MOTIVATION* and *BENEFIT*.

FIGURE 3 – EXPERIMENT: CONCEPTUAL MODELS OF CONDITIONAL INDIRECT EFFECTS

Panel A: Moderating Effect of Red Flags in the Audit Evidence (H2)



Panel B: Moderating Effect of Auditors' Personal Histories of Rewards for Costly Skepticism

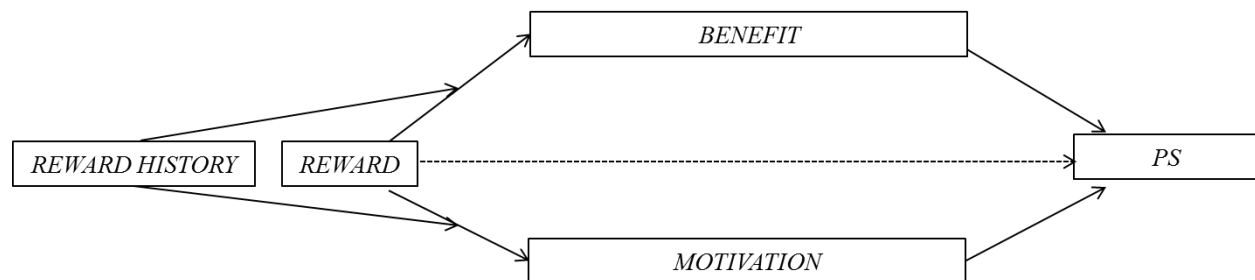


Figure 3 Panel A illustrates how the effects of both *BENEFIT* and *MOTIVATION* on *PROFESSIONAL SKEPTICISM* (*PS*) are predicted in H2 to be moderated by *RED FLAG*. *REWARD* is manipulated between participants as an interim evaluation of “Met Expectations” (No Reward) or “Exceeded Expectations” (Reward). *BENEFIT* is a self-assessment of how participants would rank in their class (compared to other seniors in their office) from receiving the evaluation in *REWARD*, collected on a scale ranging from 1 (Bottom of My Class) to 10 (Top of My Class). *MOTIVATION* is a self-assessment of how motivated participants were to exercise professional skepticism during the task. Participants responded on a scale ranging from 1 (Not at All Motivated) to 10 (Extremely Motivated). *PROFESSIONAL SKEPTICISM* (*PS*) is coded as “1” if participants concluded that more work would be required in the task and “0” if participants concluded the balance appeared reasonable. *RED FLAG* was manipulated at three levels between participants as MINOR, MODERATE, and SEVERE. In all conditions, sales growth was constant at positive 9%. In the MINOR condition, NFM growth was -1%. In the MODERATE and SEVERE conditions, NFM growth was -21% and -41%, respectively. Figure 3 Panel B depicts how the effects of *REWARD* on both *BENEFIT* and *MOTIVATION* are expected to be moderated by *REWARD HISTORY*. *REWARD HISTORY* is a self-assessment of the likelihood that managers in the participant’s own personal experience would reward costly skepticism, collected on a scale ranging from 1 (Would Not Reward Skepticism) to 10 (Would Definitely Reward Skepticism). For analyses, we partition into HIGH (top third) and LOW (bottom two thirds).



TABLE 1 – EXPERIMENT: DESCRIPTIVE STATISTICS

Panel A – Descriptive Statistics by *REWARD* for *BENEFIT*, *MOTIVATION*, & *PS*

|                         | <u>MEDIATORS</u>          |                           | <u>OUTCOME MEASURE</u>    |
|-------------------------|---------------------------|---------------------------|---------------------------|
|                         | <i>BENEFIT</i>            | <i>MOTIVATION</i>         | <i>PS</i>                 |
| <b><i>NO REWARD</i></b> | 4.90<br>(1.59)<br>n = 53  | 7.43<br>(2.01)<br>n = 53  | 0.53<br>(0.50)<br>n = 53  |
| <b><i>REWARD</i></b>    | 8.65<br>(1.63)<br>n = 57  | 7.56<br>(2.12)<br>n = 57  | 0.47<br>(0.50)<br>n = 57  |
| <b><i>MEAN</i></b>      | 6.84<br>(2.47)<br>n = 110 | 7.50<br>(2.06)<br>n = 110 | 0.50<br>(0.50)<br>n = 110 |

Panel B – Cell Means by *REWARD* & *RED FLAG* for *PS*

|                  | <i>RED FLAG</i> |                 |               |             |
|------------------|-----------------|-----------------|---------------|-------------|
|                  | <i>MINOR</i>    | <i>MODERATE</i> | <i>SEVERE</i> | <i>MEAN</i> |
| <i>NO REWARD</i> | 0.35            | 0.50            | 0.70          | 0.53        |
|                  | n = 17          | n = 16          | n = 20        | n = 53      |
| <i>REWARD</i>    | 0.45            | 0.53            | 0.44          | 0.47        |
|                  | n = 20          | n = 19          | n = 18        | n = 57      |
| <i>MEAN</i>      | 0.41            | 0.51            | 0.58          | 0.50        |
|                  | n = 37          | n = 35          | n = 38        | n = 110     |

Table 1 presents descriptive statistics. Panels A & B present descriptive statistics for *BENEFIT*, *MOTIVATION*, and *PROFESSIONAL SKEPTICISM (PS)* across the experimental conditions of *REWARD* and *RED FLAG*. *REWARD* is manipulated between participants as an interim evaluation of “Met Expectations” (No Reward) or “Exceeded Expectations” (Reward). *BENEFIT* is a self-assessment of how participants would rank in their class (compared to other seniors in their office) from receiving the evaluation in *REWARD*, collected on a scale ranging from 1 (Bottom of My Class) to 10 (Top of My Class). *MOTIVATION* is a self-assessment of how motivated participants were to exercise professional skepticism during the task. Participants responded on a scale ranging from 1 (Not at All Motivated) to 10 (Extremely Motivated). *PROFESSIONAL SKEPTICISM (PS)* is coded as “1” if participants concluded that more work would be required in the task and “0” if participants concluded the balance appeared reasonable. *RED FLAG* was manipulated at three levels between participants as MINOR, MODERATE, and SEVERE. In all conditions, sales growth was constant at positive 9%. In the MINOR condition, NFM growth was -1%. In the MODERATE and SEVERE conditions, NFM growth was -21% and -41%, respectively.

TABLE 2 – EXPERIMENT

Panel A: Hypotheses 1A &amp; 1B – Direct and Indirect Effects of Rewards for Costly Skepticism

|   | MEDIATORS             |                           | OUTCOME            |    |
|---|-----------------------|---------------------------|--------------------|----|
|   | <i>BENEFIT</i><br>(i) | <i>MOTIVATION</i><br>(ii) | <i>PS</i><br>(iii) |    |
| <i>CONSTANT</i>   | 4.99 ***<br>(21.94)   | 7.43 ***<br>(26.14)       | 0.26<br>(0.29)     |    |
| <i>REWARD</i>   | 3.66 ***<br>(11.58)   | 0.13<br>(0.32)            | 1.08<br>(1.69)     | *  |
| <i>BENEFIT</i>  |                       |                           | -0.34<br>(-2.52)   | ** |
| <i>MOTIVATION</i>   |                       |                           | 0.20<br>(1.85)     | ** |
| <i>N</i>  | 110                   | 110                       | 110                |    |
| <b>Confidence Intervals</b>                                   |                       |                           |                    |    |
| Indirect Effect <i>REWARD</i> → <i>BENEFIT</i> → <i>PS</i>    |                       |                           | (-2.08, -0.44)     |    |
| Indirect Effect <i>REWARD</i> → <i>MOTIVATION</i> → <i>PS</i> |                       |                           | (-0.09, 0.25)      |    |

Panel B: Hypothesis 2 – Moderating Effect of Red Flag Severity

| MEDIATOR          | RED FLAG              |                          |                        |
|-------------------|-----------------------|--------------------------|------------------------|
|                   | <i>MINOR RED FLAG</i> | <i>MODERATE RED FLAG</i> | <i>SEVERE RED FLAG</i> |
| <i>BENEFIT</i>    | (-2.39, 2.16)         | (-3.23, 0.23)            | (-4.62, -0.53)         |
| <i>MOTIVATION</i> | (-0.91, 0.09)         | (-0.20, 0.72)            | (-1.05, 0.62)          |

\*\*\* p &lt; 0.01, \*\* p &lt; 0.05, \* p &lt; 0.10

Panel A of Table 2 presents standardized coefficients from results of regressions for our tests of H1A and H1B. *REWARD* is manipulated between participants as an interim evaluation of “Met Expectations” (No Reward) or “Exceeded Expectations” (Reward). *BENEFIT* is a self-assessment of how participants would rank in their class (compared to other seniors in their office) from receiving the evaluation in *REWARD*, collected on a scale ranging from 1 (Bottom of My Class) to 10 (Top of My Class). *MOTIVATION* is a self-assessment of how motivated participants were to exercise professional skepticism during the task. Participants responded on a scale ranging from 1 (Not at All Motivated) to 10 (Extremely Motivated). *PROFESSIONAL SKEPTICISM (PS)* is coded as “1” if participants concluded that more work would be required in the task and “0” if participants concluded the balance appeared reasonable. The dependent measure is indicated in the column heading. *BENEFIT* and *MOTIVATION* are continuous measures, thus the coefficients are standardized OLS coefficients and test statistics are t-scores. *PROFESSIONAL SKEPTICISM (PS)* is a dichotomous measure, thus the coefficients are standardized logit coefficients and test statistics are Z-scores. Confidence intervals are 90% bias-corrected confidence intervals, and significance of the indirect effect is indicated by confidence intervals that do not include zero.

## TABLE 2 – EXPERIMENT (continued)

Panel B of Table 2 presents confidence intervals for the indirect effects of *BENEFIT* and *MOTIVATION* on *PS* at each level of the *RED FLAG (RF)* manipulation. *RED FLAG* was manipulated at three levels between participants as MINOR, MODERATE, and SEVERE. In all conditions, sales growth was constant at positive 9%. In the MINOR condition, NFM growth was -1%. In the MODERATE and SEVERE conditions, NFM growth was -21% and -41%, respectively. Support for H2 would be indicated by (1) decreasing significance of the *BENEFIT* indirect effect as the *RF* becomes more severe and/or (2) increasing significance of the *MOTIVATION* indirect effect as the *RF* becomes more severe. Neither condition is supported.

TABLE 3 – EXPERIMENT: RED FLAG RECOGNITION & COMMUNICATION TO MANAGER

Cell Means for *RECOGNITION* and *INFORM MANAGER*

|                  | <i>RED FLAG</i> | <i>RECOGNITION</i> | <i>INFORM MANAGER</i> |
|------------------|-----------------|--------------------|-----------------------|
| <i>NO REWARD</i> | <i>MINOR</i>    | 4.44               | 0.31                  |
|                  | <i>MODERATE</i> | 6.53               | 0.36                  |
|                  | <i>SEVERE</i>   | 7.90               | 0.79                  |
| <i>REWARD</i>    | <i>MINOR</i>    | 5.75               | 0.44                  |
|                  | <i>MODERATE</i> | 7.74               | 0.50                  |
|                  | <i>SEVERE</i>   | 7.22               | 0.53                  |

Table 3 depicts the cell means for *RECOGNITION* and *INFORM MANAGER* across the experimental conditions of *REWARD* and *RED FLAG*. *RECOGNITION* is the participant's assessment of the extent to which the NFM trend is inconsistent with the growth in recorded sales, collected on a scale ranging from 1 (Very Small) to 10 (Very Large). *INFORM MANAGER* is coded as "1" if participants explicitly wrote down the NFM trend in the list of items they would communicate to their manager and "0" otherwise. *RED FLAG* was manipulated at three levels between participants as MINOR, MODERATE, and SEVERE. In all conditions, sales growth was constant at positive 9%. In the MINOR condition, NFM growth was -1%. In the MODERATE and SEVERE conditions, NFM growth was -21% and -41%, respectively. *REWARD* is manipulated between participants as an interim evaluation of "Met Expectations" (No Reward) or "Exceeded Expectations" (Reward).

TABLE 4 – EXPERIMENT: AUDITORS’ PERSONAL EXPERIENCE WITH REWARDS FOR COSTLY SKEPTICISM

Panel A: Cell Means

|                  | <i>REWARD HISTORY</i> | <i>BENEFIT</i> | <i>MOTIVATION</i> | <i>PS</i> |
|------------------|-----------------------|----------------|-------------------|-----------|
|                  |                       |                |                   |           |
| <i>NO REWARD</i> | <i>NEGATIVE</i>       | 4.90           | 6.83              | 0.49      |
|                  | <i>POSITIVE</i>       | 5.18           | 8.53              | 0.59      |
| <i>REWARD</i>    | <i>NEGATIVE</i>       | 8.54           | 7.46              | 0.43      |
|                  | <i>POSITIVE</i>       | 8.85           | 7.75              | 0.55      |

Panel B: Regression for Conditional Indirect Effect

|                              | <i>BENEFIT</i>      | <i>MOTIVATION</i>   | <i>PS</i>           |  |
|------------------------------|---------------------|---------------------|---------------------|--|
|                              | (i)                 | (ii)                | (iii)               |  |
| <i>CONSTANT</i>              | 4.90 ***<br>(17.32) | 6.83 ***<br>(20.15) | 0.21<br>(0.23)      |  |
| <i>REWARD</i>                | 3.64 ***<br>(9.23)  | 0.63<br>(1.33)      | 1.05<br>(1.64)      |  |
| <i>REWARD HISTORY</i>        | 0.28<br>(0.56)      | 1.70 **<br>(2.86)   |                     |  |
| <i>REWARD*REWARD HISTORY</i> | 0.03<br>(0.05)      | -1.41 *<br>(-1.73)  |                     |  |
| <i>MOTIVATION</i>            |                     |                     | 0.22 **<br>(1.98)   |  |
| <i>BENEFIT</i>               |                     |                     | -0.35 **<br>(-2.56) |  |
| <i>N</i>                     | 105                 | 105                 | 105                 |  |

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ . Panel A of Table 4 depicts descriptive statistics for the *BENEFIT*, *MOTIVATION*, and *PROFESSIONAL SKEPTICISM (PS)* across levels of the *REWARD* and *REWARD HISTORY* variables. *BENEFIT* is a self-assessment of how participants would rank in their class (compared to other seniors in their office) from receiving the evaluation in *REWARD*, collected on a scale ranging from 1 (Bottom of My Class) to 10 (Top of My Class). *MOTIVATION* is a self-assessment of how motivated participants were to exercise professional skepticism during the task. Participants responded on a scale ranging from 1 (Not at All Motivated) to 10 (Extremely Motivated). *PROFESSIONAL SKEPTICISM (PS)* is coded as “1” if participants concluded that more work would be required in the task and “0” if participants concluded the balance appeared reasonable. *REWARD* is manipulated between participants as an interim evaluation of “Met Expectations” (No Reward) or “Exceeded Expectations” (Reward). *REWARD HISTORY* is a self-assessment of the likelihood that managers in the participant’s own personal experience would reward costly skepticism, collected on a scale ranging from 1 (Would Not Reward Skepticism) to 10 (Would Definitely Reward Skepticism). For analyses, we partition into HIGH (top third) and LOW (bottom two thirds).

TABLE 4 – EXPERIMENT: AUDITORS’ PERSONAL EXPERIENCE WITH REWARDS  
FOR COSTLY SKEPTICISM (continued)

Panel B of Table 4 presents standardized coefficients from results of regressions for our supplemental analyses. The dependent measure is indicated in the column heading. *BENEFIT* and *MOTIVATION* are continuous measures, thus the coefficients are OLS coefficients and test statistics are t-scores. *PS* is a dichotomous measure, thus the coefficients are logit coefficients and test statistics are Z-scores.

TABLE 5 – SURVEY: DETERMINANTS OF REWARDS FOR COSTLY SKEPTICISM

|                   |                           | <i>EVAL</i><br>(i) |     | <i>EVAL</i><br>(ii) |    | <i>EVAL</i><br>(iii) |    | <i>EVAL</i><br>(iv) |     |
|-------------------|---------------------------|--------------------|-----|---------------------|----|----------------------|----|---------------------|-----|
| <u>Incentives</u> | <i>AFFECT OWN EVAL</i>    | 0.37               | *** |                     |    |                      |    | 0.36                | *** |
|                   | <i>GENERAL PS REWARDS</i> | 0.02               |     |                     |    |                      |    | -0.16               |     |
|                   | <i>REWARD HISTORY</i>     | 0.04               |     |                     |    |                      |    | 0.04                |     |
| <u>Traits</u>     | <i>TRAIT SKEPTICISM</i>   |                    |     | < 0.01              |    |                      |    | 0.07                |     |
|                   | <i>MY RANK</i>            |                    |     | 0.19                | ** |                      |    | 0.08                |     |
| <u>Knowledge</u>  | <i>CONSULTATIVE</i>       |                    |     |                     |    | 0.21                 | ** | 0.26                | **  |
|                   | <i>TOTAL EXPERIENCE</i>   |                    |     |                     |    | 0.19                 | *  | 0.20                | **  |
|                   | <i>INDUSTRY</i>           |                    |     |                     |    | -0.11                |    | -0.15               |     |
|                   | <i>NFM</i>                |                    |     |                     |    | 0.17                 | *  | 0.17                | *   |
|                   | <i>FRAUD TRAINING</i>     |                    |     |                     |    | -0.06                |    | -0.05               |     |
|                   | <i>MM BASERATE</i>        |                    |     |                     |    | -0.10                |    | -0.08               |     |
|                   | R-squared                 | 0.14               |     | 0.04                |    | 0.09                 |    | 0.23                |     |
|                   | N                         | 127                |     | 127                 |    | 127                  |    | 127                 |     |

\*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.10.

Table 5 depicts coefficients from OLS regressions. The dependent measure is *EVAL*, which is the participant's assessment of the junior auditor's performance, collected on a scale ranging from -5 (Below Expectations) to 0 (Met Expectations) to +5 (Above Expectations). *AFFECT OWN EVAL* is an assessment of how the partner will evaluate the participant's own performance collected on a scale ranging from -5 (Below Expectations) to 0 (Met Expectations) to +5 (Above Expectations). *GENERAL PS REWARDS* is the self-reported experience being rewarded by managers for any skepticism, collected on a scale ranging from 1 (Would Not Reward Skepticism) to 10 (Would Definitely Reward Skepticism). *REWARD HISTORY* is a self-assessment of the likelihood that managers in the participant's own personal experience would reward costly skepticism, collected on a scale ranging from 1 (Would Not Reward Skepticism) to 10 (Would Definitely Reward Skepticism). For analyses, we partition into HIGH (top third) and LOW (bottom two thirds). *TRAIT SKEPTICISM* is measured via the Hurtt (2010) scale. *MY RANK* is a self-assessment of how the participant currently rank in their class, collected on a scale ranging from 1 (Bottom of My Class) to 10 (Top of My Class). *CONSULTATIVE* is the extent to which participants consult their supervisors while exercising skepticism, collected on a scale ranging from 0 (Never) to 10 (Always). *TOTAL EXPERIENCE* is the participant's audit experience in years & months (converted to decimals). *INDUSTRY* is the participant's percentage of chargeable hours in manufacturing during the past three years (the industry in our case). *NFM* is the percentage of substantive analytical procedures conducted by the participant that have involved NFMs. *FRAUD TRAINING* is the participant's number of hours of training on financial statement fraud. *MM BASERATE* is the participant's estimate of the percentage of pre-audit financial statements that contain a material misstatement.